

Dr hab. Bartosz Wilczyński
Wydział Matematyki, Informatyki i Mechaniki
Uniwersytet Warszawski
ul. Banacha 2, 02-089 Warszawa
b.wilczynski@mimuw.edu.pl

Warszawa, 20 .05.2020 r.

Recenzja rozprawy doktorskiej

Tomasza Prejzenda

zatytułowanej:

*Wykorzystanie kontrolowanych języków naturalnych
do modelowania systemów dynamicznych w bioinformatyce*

1. Problem badawczy i jego znaczenie

Konstrukcja modeli matematycznych, czy szerzej, obliczeniowych, jest obecnie ważnym elementem badań nad zjawiskami przyrodniczymi. W ostatnich dziesięcioleciach, w związku z ogromnym wzrostem znaczenia komputerów zarówno w pozyskiwaniu jak i analizie danych eksperymentalnych, wiele badań w szeroko pojętej dziedzinie nauk przyrodniczych nie byłoby możliwe bez modelowania matematycznego. Dotyczy to zarówno badań w dziedzinach już klasycznych, takich jak ekologia, czy epidemiologia, jak i najnowszych badań opartych o badania molekularne, takie jak modelowanie i analiza regulacji transkrypcji genów. Bardzo częstym przedmiotem modelowania matematycznego są procesy dynamiczne, zachodzące w czasie (i/lub przestrzeni), gdzie ewolucja w czasie zestawu zmiennych ilościowych może być reprezentowana jako układ równań różniczkowych

Praca doktorska mgra Tomasza Prejzenda podejmuje temat wykorzystania kontrolowanych języków naturalnych do modelowania układów dynamicznych w bioinformatyce i jako przykłady takich układów dynamicznych wykorzystuje właśnie modele związane z systemem CRISPR/Cas9 u bakterii, a także modele drapieżnik-ofiara. Jako, że rozprawa dotyczy głównie technologii modelowania, dobór modeli nie ma tak dużego znaczenia, ale trzeba przyznać, że zwłaszcza tematyka CRISPR/Cas9 jest w ostatnich latach bardzo aktywnie badanym zjawiskiem i w związku z tym wykorzystanie tego przykładu zwiększa aktualność rozprawy i jej wyników.

2. Wkład autora

Recenzowana tu rozprawa jest bardzo obszerna, składa się z ośmiu rozdziałów (nie licząc „Podsumowania”) oraz 4 dodatków. Nawet w streszczeniu autor podaje 5 różnych „najważniejszych” celów do osiągnięcia w swojej rozprawie, począwszy od klasyfikacji dostępnych metod modelowania (1), poprzez projekt i rozwój nowego języka Modelang (2), weryfikację jego funkcjonalności (3), a na koniec budowę środowisk do analizy modeli (4), ich przechowywania i współdzielenia (5). Niestety, nie wszystkie te elementy można jednoznacznie przypisać autorowi rozprawy, choćby z powodu tego, że język Modelang powstał już wcześniej jako element prac prowadzonych w doktoracie Doktora

Wąsika. Niemniej jednak, można stwierdzić, że mgr Prejzendanc brał udział w pracach nad językiem Modelang, jako że jest współautorem głównej pracy na ten temat (Wąsik, Prejzendanc & Błażewicz, 2013). Pozostałe części pracy autora tej dysertacji były opublikowane także jako prace wieloautorskie: zarówno jeśli chodzi o klasyfikację modeli (Prejzendanc, Wąsik & Błażewicz, 2016), jak i środowisko do przechowywania modeli infekcji wirusowych VirDB (Wąsik et al, 2019). Niewątpliwie autorskim dziełem doktoranta pozostaje cel trzeci, czyli porównanie funkcjonalności modeli w języku Modelang z innymi rozwiązaniami (pakietem Mathematica i CellDesignerem). Z mojego punktu widzenia, te właśnie wyniki, według mojej obecnej wiedzy jeszcze nie opublikowane, stanowią główny trzon autorskiego wkładu magistra Prejzendanca w analizowaną tu rozprawę.

W związku z tym, że ustawa o stopniach naukowych i tytule naukowym nakłada na recenzentów obowiązek stwierdzenia czy autorski wkład doktoranta w prace przedstawione w rozprawie jest wystarczający do nadania stopnia doktora, brak jednoznacznego określenia wkładu doktoranta w wyniki uzyskane w poszczególnych celach, czy zadaniach rozwiązywanych w doktoracie jest niewątpliwym utrudnieniem w ocenie tej pracy i może powodować zaniżoną ocenę wkładu magistra Prejzendanca w przedstawione wyniki. Niemniej trzeba przyznać, że nawet w tych zadaniach, w których nie sposób jednoznacznie określić wkładu autora w koncepcje rozwiązań, niewątpliwie oceniana tu rozprawa stanowi rozszerzenie treści w stosunku do opublikowanych prac, choćby ze względu na swoją objętość, znacznie przekraczającą sumaryczną objętość opublikowanych artykułów. Można zatem stwierdzić, że wkład autora w zadanie weryfikacji modeli jest kluczowy, natomiast jego wkład w pozostałe 4 osiągnięte cele polega co najmniej na uporządkowaniu pojęć i ich rozwinięciu w przedstawionej rozprawie.

3. Poprawność

Jeśli chodzi o poprawność przedstawionych wyników, to, jako że praca zawiera głównie implementacje pewnych rozwiązań w postaci programów komputerowych oraz zapisy modeli matematycznych w języku Modelang, nie sposób definitywnie określić poprawności lub niepoprawności rozumowań, tak jak w przypadku prac poświęconych matematycznym modelom tych samych zjawisk.

Można natomiast przyrzeć się przedstawionym w doktoracie nowym reprezentacjom opublikowanych wcześniej modeli matematycznych i porównać ich przedstawione działanie w języku Modelang do wcześniej znanych implementacji w innych językach. Tutaj najlepiej oceniam część poświęconą modelom infekcyjnym (HCV i HIV), które zasadniczo są opisane zgodnie z literaturą i nie mam do nich zastrzeżeń. Niestety, w części poświęconej modelom drapieżnik-ofiara, oraz w części poświęconej modelowi CRISPR-Cas9, autor nie ustrzegł się pewnych błędów.

W sekcji 6.2.1.1, gdzie autor opisuje „uproszczony model Lotki-Volterra”, autor wspomina o tym, że „przy nieobecności drapieżników, roślinożercy mogą [...] się rozwijać w sposób niepohamowany”. Jest to prawda w przypadku najprostszego modelu Lotki-Volterra, jednak we wzorach przypisanych do tego modelu autor uwzględnia już składnik $-bY^2$, który właśnie zapobiega temu efektowi poprzez „karanie” populacji roślinożerców za ich nadmierny rozrost. Nie jest jasne, czy autor chciał wpisać tu wzór prostszy (bez składnika $-bY^2$), czy też niepotrzebna jest dyskusja „niepohamowanego rozwoju” roślinożerców i jego związku z modelem Rosenzweiga-McArthura. Niemniej, jest to niedociągnięcie świadczące co najmniej o pośpiechu przy spisywaniu rozprawy, albo o niedostatecznym opanowaniu tej tematyki przez doktoranta.

W rozdziale 7, gdzie autor opisuje model systemu CRISPR/Cas9 problem jest moim zdaniem nieco poważniejszy, choć można przypisać go bardziej do dziedziny modelowanej, w tym przypadku biologii, co zmniejsza jego wpływ na ocenę pracy doktoranta. Mianowicie, głównym celem rozdziału 7, pod tytułem „Model mechanizmu CRISPR/Cas9” jest reprezentacja modelu z pracy Iranzo et al, z

roku 2013, w której zaproponowano model działania systemu CRISPR/Cas9 w replikacji bakterii. Oryginalny model zawiera 3 równania różniczkowe, podane w rozprawie jako Równanie 12, na stronie 157. Niestety, ten model nie został ostatecznie zaimplementowany w języku Modelang, a jedynie „model uproszczony” składający się z dwóch równań różniczkowych (Równanie 13, strona 158). Jest to w pracy uzasadnione w ten sposób, że „pełny model mechanizmu CRISPR/Cas9[...] zawiera dużą liczbę parametrów, których uwzględnienie[...] komplikuje prace badawcze”. Niestety, ten uproszczony model nie zawiera zmiennych N_{b+} i N_{b-} , czyli rozróżnienia na bakterie niosące wirusa i pozostałe. Podsumowując, uproszczony model rozważany w pracy może stanowić interesujący przykład użycia dla systemu Modelang, ale argumentowanie, że w jakiś sposób jest to model „mechanizmu CRISPR/Cas” jest stwierdzeniem nie mającym podstaw w rzeczywistości. Z tego względu, mimo że sam model w języku Modelang stanowi autorski wkład doktoranta, jego opis niezgodny z rzeczywistością, istotnie obniża wartość tego wyniku.

Także opis eksperymentalnego porównania modelu wygenerowanego z opisu w języku Modelang z układem równań różniczkowych (w sekcji 7.3.2.2) pozostawia niestety wiele do życzenia. Autor w zasadzie ogranicza się do stwierdzenia, że „Przeprowadzona analiza wykazała zbieżność danych”, jednak widzimy tylko wykres uzyskany przy pomocy programu Mathematica, a „zbieżność danych” nie jest poparta żadnymi wskaźnikami liczbowymi. W tej sprawie poprosiłem o wyjaśnienia autora, który udzielił mi dostępu do danych z tego eksperymentu. Dane te rzeczywiście wykazują zbieżność na poziomie dokładności numerycznej komputera, jednak brak tego opisu w rozprawie stanowi pewną usterkę.

4. Wiedza kandydata

Kandydat zawarł w swej rozprawie bardzo obszerne części, które stanowią uporządkowanie istniejącego stanu wiedzy. Całe rozdziały 2, 3 i 4 (zajmujące około 60 stron tekstu) stanowią przegląd podstaw biologicznych, matematycznych i informatycznych przedstawionych dalej prac. Także dalsze rozdziały zawierają liczne odniesienia do wcześniejszej literatury dotyczącej pierwotnych implementacji rozważanych modeli w rozdziałach 6 i 7. Większość tych opisów stanowi dobre podsumowanie istniejącej wiedzy w kilku pokrewnych dziedzinach modelowania matematycznego i jest adekwatnie udokumentowana odniesieniami do literatury. Spis literatury zawiera kilkadziesiąt pozycji i są to w znakomitej większości prace istotnie powiązane z tematem.

Pewne niedociągnięcia stylistyczne, w rodzaju sformułowań potocznych (np. „zamodelowane”, czy „całkowicie intuicyjny model”) czy kalek z języka angielskiego (np. „framawork”) nie stanowią w pracy tak obszernej tematyce poważnego problemu. Dziedziny techniczne niewątpliwie stawiają autora rozprawy pisaną w języku polskim przed trudnymi często tłumaczeniami fachowych terminów i tutaj trzeba oddać mgrowi Prejzendorowi, że wybrał z większości tych problemów nie gorzej niż większość autorów prac w bioinformatyce.

Można też, niestety, dopatrzeć się pewnych niedociągnięć w opisie metod informatycznych, zwłaszcza w odniesieniu do języków formalnych i kompilatorów (Rozdziały 3.5 i 3.6). Uważam, że po prostu ta część jest niepotrzebna i niemożliwa do wykonania w tak krótkim tekście. Np. rozdział 3.5 poświęcony językom formalnym zajmuje niecałe 3 strony i próbuje opisać dużą część tematyki języków formalnych od automatów skończonych do Maszyn Turinga. Wydaje mi się, że aby to zrobić dobrze, konieczne byłoby go znaczne rozwinięcie, pewnie o kolejne 10 lub więcej stron, a nie ma to bezpośredniego związku z rozprawą – która jedynie stosuje już gotowe pakiety do kompilacji programów w Modelangu. Podsumowując – mimo pewnych niedociągnięć stylistycznych – część opisująca istniejący stan wiedzy stanowi moim zdaniem najlepszą część rozprawy i świadczy o tym, że kandydat poświęcił sporo czasu na poznanie różnych metod związanych z rozległą dziedziną modelowania.

5. Inne uwagi

Magister Prejzendanc jest współautorem 3 prac w recenzowanych czasopismach międzynarodowych. Wszystkie one wiążą się z częściami rozprawy co, mimo że są to prace wieloautorskie, stanowi jego wkład w rozwój dziedziny. Mimo niewątpliwych błędów i niedociągnięć w nieopublikowanej części poświęconej modelowaniu, rozprawa ta świadczy o zgromadzonej przez autora wiedzy i może (po poprawieniu wspominanych przeze mnie usterek) stanowić podstawę do dalszych publikacji.

6. Podsumowanie

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez artykuł 13 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym (z późniejszymi zmianami)¹ moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

A. Czy rozprawa zawiera oryginalne rozwiązanie problem naukowego? (wybierz jedną opcję stawiając znak X)

☐	☒	☐	☐	☐
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

B. Czy po przeczytaniu rozprawy zgadzasz się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka lub Automatyka i Robotyka?

☒	☐	☐	☐	☐
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

C. Czy kandydat umiejętność samodzielnego prowadzenia pracy naukowej?

☐	☐	☒	☐	☐
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

Bartosz Wilczyński

¹ http://www.nauka.gov.pl/g2/oryginal/2013_05/b26ba540a5785d48bee41aec63403b2c.pdf